DOCUMENT RESUME

ED 137 378 TM 006 183

AUTHOR Carroll, C. Dennis; McKeown, Robin J.
TITLE MANOVA for Longitudinal Designs with Multiple Dependent Variables.
PUB DATE [Apr 77]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Analysis of Variance; Hypothesis Testing; *Longitudinal Studies; *Statistical Analysis; Statistical Bias; Tests of Significance
IDENTIFIERS *Multivariate Analysis of Variance

ABSTRACT

The application of a multivariate analytic technique for the analysis of data from longitudinal designs with multiple dependent variables is presented. The technique is the multivariate generalization of univariate repeated measures ANOVA. An application of the technique to data collected using materials from the Asian Studies Curriculum Project is included. The example analysis indicated the technique is viable and should be a useful tool for the methodologist/evaluator. (Author)

AERA Session #27.14

MANOVA FOR LONGITUDINAL DESIGNS WITH MULTIPLE DEPENDENT VARIABLES

by
C. Dennis Carroll
U.S. Office of Education
and
Robin J. McKeown
University of California, Riverside

Longitudinal designs are frequently employed in educational research and evaluation. Studies including pre-post-retention, post-retention, or pre-post testing are longitudinal studies. The simultaneous utilization of multiple experimental treatments or experimental and control treatments in a longitudinal study automatically yields some form of a longitudinal design. When longitudinal designs are employed in conjunction with multiple dependent variables, the appropriate analytic technique may be a subject for debate. The purpose of this paper is to describe an analytic technique that removes or minimizes the difficulties encountered with many of the currently advocated techniques. In addition, an application of the proposed technique to data collected using materials from the Asian Studies Curriculum Project is presented.

CURRENT TECHNIQUES

Technique I

An analysis technique frequently advocated for longitudinal designs with multiple dependent variables is multivariate analysis of variance (MANOVA) with covariate adjustments. For designs including pre-post-retention tests, separate analyses of the post and retention tests are required following adjustment for the pretest scores. There are at least two methods of employing the pretest scores as covariates. That is, for dependent variable set

$$Y_1, Y_2, Y_3, \ldots, Y_i, \ldots, Y_k$$

it is possible to covary the pretest Y_1 scores from the post test Y_1 scores, the pretest Y_2 scores from the post test Y_2 scores, the pretest Y_3 scores from the post test Y_3 scores, etc. It is also possible to use the entire set (Y_1 through Y_k) of pretest scores as covariates for each post test score (e.g., Y_i). Finally, the observations of Elashoff (1969) with respect to ANOCOVA may be used to question these procedures. That is, heterogeneity of regression or selection bias may generate very misleading results.

The MANOVA with covariance adjustments technique suffers from many problems. A list of these problems includes the following:

(1) Exponentiation of p-values for designs requiring separate analyses. This problem is inherent with any design including post and retention tests.

(2) The frequently required separate analyses eliminate the Gestalt from the analysis. In addition, the use of separate analyses assumes some form of interaction (e.g., differential trend) is present in the system. Interaction questions should be answered empirically, rather than by assumption.

(3) The techniques of covariate adjustment are not well-defined. That is, with multiple dependent variables there is no single field accepted technique of reducing the error using the covariates.

(4) The frequency of heterogeneity of regression in educational research and evaluation as well as the frequency of selection biases in educational evaluation suggests this technique may have questionable utility for many problems.

(5) No trend or differential trend analyses are incorporated in the technique. Concern for change, especially in educational evaluation, suggests this technique ignores a very important issue.

Technique II

Analysis techniques frequently employed in evaluation studies are variations of the separate analysis technique. This approach can take three forms. The first approach consists of separate MANOVAs for each test occasion. That is, separate analyses are conducted for the pre-post- and/or retention tests. The major problems associated with the use of this approach are (#1), (#2), and (#5) as listed for Technique I.

The second approach is to employ separate univariate repeated measures ANOVAs for each dependent variable. The problem associated with this technique is similar to the exponentiation of p-values problem. In general, the failure to recognize the covariance structure in a dependent variable set is the source of the difficulties. The univariate p-values are accurate only when the dependent variables are pairwise orthogonal. Multivariate adjustments to the F-statistics may be used to correct for bias in the p-values (Harris, 1975) but these are rarely used.

The third approach is to employ separate MANOVAs for each dependent variable treating each test occasion observations as distinct dependent variables for the analyses. Profile analyses, which yield hypotheses sets similar to univariate repeated measures ANOVA, are frequently employed with this technique. The problems associated with this technique are a result of its failure to account for the covariance structure in the dependent variable set. While this technique minimizes the unknown difficulties with one of the two covariance assumptions of univariate repeated measures ANOVA, it suffers from the lack of p-value adjustment procedures for the covariance structure of the dependent variable set. The degree of this problem is unknown, but the capitalization upon chance associated with MANOVA may increase the degree of the problem in relation to univariate techniques.

Technique III

A less frequently employed technique consists of applying MANOVA to an extended dependent variable set. That is, the pre-post-retention observations of each dependent variable are considered to be distinct dependent variables for the analysis. Using a pre-post-retention design results in three times the original number of dependent variables in the extended dependent variable set.

While this is the only technique listed thus far that eliminates or minimizes problems (#1), (#2), (#3), and (#4) as listed for Technique I, this technique (III) is not free of problems. Specifically, problem (#5) as listed for Technique I is a telling source of difficulty. In addition, an interpretational problem may be experienced with this technique. That is, it is possible for the MANOVA generated composite to be a difference or sum of pretest observations of one dependent variable and posttest observations of another dependent variable. To interpret composites of this form frequently presents a great deal of difficulty.

Proposed Technique

The appropriate technique should take into account the following necessary attributes:

(1) The technique includes a complete hypothesis set, and thereby gives the analysis an appropriate Gestalt. Included in the hypothesis set are issues of trend and differential trends.

(2) The technique produces p-values after adjusting for the covariance structure in the dependent variable set and the covariance structure in the test occasions.

(3) The technique is well-defined in terms of representing a single procedure.

(4) Only the generalizability of the results are threatened by selection biases. Specifically, the p-values reflect unbiased estimates of the "population" generalized from the sample employed.

(5) The results are interpretable in terms of the dependent variable set only. That is, the composites produced are not confounded with test occasions or adjustments for specific test occasions.

The analytic procedure advocated by this paper is the multivariate generalization of univariate repeated measures ANOVA. Simple univariate repeated measures ANOVA identifies the effects and error associated with distinct groups of students (usually referred to as "between") by summing over the repeated facet. The remaining effects and error (referred to as "within") are associated with hypotheses concerning trends and differential trends. The multivariate generalization of repeated measures ANOVA isolates the between and within terms for multiple dependent variables.

The generalization from univariate to multivariate analysis, as described by Harris (1975), involves

> "a search for that linear combination of the various outcome variables which makes the univariate F-ratio (computed on that single, combined variable) for a particular main effect or interaction as large as possible (page 15)."

All of the variance hypotheses associated with a univariate analysis may be generalized to multivariate forms by replacing the univariate dependent variable with the multivariate composite of dependent variables. That is, for a longitudinal design the repeated measures hypothesis set may be retained in the multivariate generalization. This hypothesis set is complete and specifically includes hypotheses for trend and differential trends. In addition, the composites generated by the multivariate procedure are in terms of the dependent variable set only.

The multivariate significance tests determine p-values that have been adjusted for the maximization (or capitalization upon chance) procedures employed to generate the composites. These maximization procedures capitalize upon the covariance structure of the dependent variable set. That is, the multivariate p-values are adjusted for the covariance structure of the dependent variable set. Furthermore, the error terms associated with repeated measures ANOVA have expected values that are functions of the covariance structure in the test occasions. As a result of this incorporation of covariance structures for expected values, the p-values are appropriately estimated. Therefore, the p-values generated by repeated measures MANOVA are expected to be appropriate for the covariance structures of the dependent variables and the test occasions.

As a result of the generalized hypothesis set associated with MANOVA, the technique is as well-defined as ANOVA. An added benefit is the ease of transition from ANOVA to MANOVA. In addition the applications of MANOVA are equivalent to the applications of ANOVA (except for the multiple dependent variable set). Therefore, selection biases are expected to affect MANOVA in the same fashion that selection biases affect ANOVA.

Computationally, the generalization from ANOVA to MANOVA is straightforward. The sums of squares in ANOVA are replaced by matrices of cross-products. That is, the sums of squares for an effect A is of the form

$$SS_A = \sum_{i \varepsilon A} (\overline{Y}_{ij} - \overline{Y}_j)^2$$

The generalization of SS_A to a matrix of cross products, H_A is of the form

$$h_{rs} = \sum_{i \varepsilon A} (\overline{Y}_{ijr} - \overline{Y}_{jr})(\overline{Y}_{ijs} - \overline{Y}_{js})$$

where

h_{rs} = r, s-th element of H_A

Y_r = the r-th dependent variable, and

Y_s = the s-th dependent variable.

All sums of squares, effects and errors, may be generalized to cross product matrices in this fashion.

The F-test of ANOVA is generalized for MANOVA to a function of the greatest characteristic roots of $H^{-1}_{error}\ H_{effect}$. There are several MANOVA functions that may be used for significance tests. The most frequently occurring are presented in Morrison (1967), known as Wilks' Lambda (Λ), and Harris (1975), known as GCR θ. Harris (1975) compared these two indices and advocated GCR θ on the basis of structure and statistical power.

AN EXAMPLE: EVALUATION OF THE ASIAN STUDIES CURRICULUM PROJECT

To illustrate the use of repreated measures MANOVA, data collected 1/ for an evaluation of materials from the Asian Studies Curriculum Project (Michaelis, 1968) were analyzed. This data reflected a condition X reading level X test occasion design for six dependent variables.

The conditions were instructional strategies labeled as open inquiry, guided inquiry, expository, and control. The control condition was distinguished from the other conditions by the topic of the materials employed. The reading materials in the open inquiry, guided inquiry, and expository conditions were identical, but these conditions incorporated differing types of questions and directions interspersed in the materials. The students subjected to the expository condition were asked to respond to recall and recognition questions. Both the open inquiry and guided inquiry subjects were asked to respond to analysis, synthesis, and evaluation questions, but only the guided inquiry subjects were allowed to benefit from model answers. The number of questions interspersed in the materials was varied in the conditions to yield approximately equal reading times.

Procedurally, the ninth grade students in twelve "regular" social studies classes were randomly assigned to condition. Five weeks, or about 25 hours, of instruction using the previously described condition materials was provided.

1/ Data were collected by Dr. Robin J. McKeown.

Reading level was operationalized by blocks of the students' California Achievement Test scores. Above average and below average blocks were based on the distribution of the students' scores. The cut point in the distribution was the score associated with the 64-th percentile.

The test occasions were post and retention observations. The post-tests were administered within a week following completion of treatment. The retention tests were administered about five weeks after the post-tests.

The dependent variables were operationalized by parallel forms of experimenter constructed 2/ measures. These measures were labeled as fact recall, concept comprehension, general synthesis skill, social problem solving, interest, and involvement. For both test occasions, the sequence of administration for these scales was the same as the above listing.

The fact recall test consisted of 30 recall items. Each item required a response of one or two words and each item was scored as right or wrong. Fifteen minutes were allowed for student responses. All responses were associated with the material employed in the inquiry and expository conditions.

The concept comprehension test consisted of six concepts included in the inquiry and expository conditions' materials. For each concept the students were required to respond with a short paragraph describing and/or explaining the meaning of the concept. Each response was scored using a [0, 5]-scale by three raters with forced agreement.

2/ The measures were constructed by Dr. Robin J. McKeown.

The general synthesis skill test contained six sets of social studies materials which were unrelated to any of the experimental materials. The respondents were required to write a brief paragraph detailing any generalizations (e.g., conclusions, ideas, themes) from the material. The students were allowed a maximum if 40 minutes to respond. Each response was scored using a [0, 5]-scale by three raters with forced agreement.

The social problem solving test consisted of six social value problems related to the experimental materials. Each problem required a response consisting of a brief paragraph analyzing the problem. The students were allowed a maximum of 40 minutes to respond. Each response was scored using a [0, 5]-scale by three raters with forced agreement. Only the quality of the analysis was scored, not the actual content.

The interest scale consisted of 10 six-point Likert bipolar items in disagree-agree format. The stems of the Likert items were constructed to identify the students' interest in continuing to study about the areas covered in the experimental materials. The total of the 10 items' scores, following reversal of negative scales, was used as the interest scale score. The internal consistency reliability of these total scores was estimated by Cronbach's Alpha to be .90.

The involvement scale consisted of 10 six-point Likert bipolar items in disagree-agree format. The stems of the Likert items were constructed to determine the degree the student wanted or wished the problems presented in the experimental materials to be solved. That is, the involvement scale constructually represented the student's commitment to a solution for the problems. The total of the 10 items' scores, following reversal of negative scales, was computed. A Cronbach's Alpha of .86 was calculated as an estimate of the involvement scale's reliability.

Results

The descriptive indices computed from the data are displayed in Table 1. The data were analyzed using a 4 x 2 x 2 repeated measures MANCOVA. The components evaluated using this procedure were the effects of the three independent variables, three first-order interactions, and a second-order interaction. Each component was associated with effect and error matrices that produced a maximizing linear composite of dependent variables. The magnitude of variation in mean composite scores associated with each effect was compared to chance fluctuation using greatest characteristic root (Harris, 1975) and Wilks' Lambda (Morrison, 1963) criteria. Generation of effect and error matrices with probability indices was facilitated by a computer routine (Carroll and Klimowicz, 1975).

A condition X reading level X test occasion interactive effect was identified (GCR $\theta = .175$, $s = 3$, $m = 1$, $n = 145$, $p < .0001$; Wilks' $\Lambda = .775$, $F = 4.34$, $df = 18/407$, $p < .0001$). This interaction, as with all second order interactions, may be interpreted in six different ways. The source of this data, an evaluation of the Asian Studies Curriculum Project, indicated the interpretation of this interaction should reflect concern for the trends displayed by the conditions. Based upon this concern, the interaction was interpreted as evidence that differences in the conditions' trends were dependent upon the subjects' reading level. A more ANOVA-ish phrasing of this interpretation might be as follows: The condition X test occasion interactions are different for above and below average readers. Methodologically, the second order interaction provided empirical justification for separate analyses of the above average and below average subjects' data.

Analysis for Above Average Readers

The condition X test occasion data of the above average readers were submitted to a 4 x 2 repeated measures MANOVA. A condition X test occasion interactive effect was identified (GCR $\theta = .538$, $s = 3$, $m = 1$, $n = 71$, $p < .0001$; Wilks' $\Lambda = .448$, $F = 7.45$, $df = 18/407$, $p < .0001$). In a fashion similar to that employed for the second-order interaction, the condition X test occasion interaction was viewed as empirical justification for separate analyses of the post and retention test data.

Post-test analyses for above average readers

The data were submitted to a simple MANOVA. The mean composite scores of the subjects in the four conditions were different (GCR $\theta = .756$, $s = 3$, $m = 1$, $n = 71$, $p < .0001$; Wilks' $\Lambda = .103$, $F = 27.91$, $df = 18/407$, $p < .0001$). The composite associated with the effect of conditions was (fact recall + 2* concept comprehension - general synthesis skill). A general but tentative interpretation of the composite as a discrepancy between lower and higher taxonomic level skills could be posed. However, measure specific tests were also generated for theoretical clarity.

Univariate analyses of the effects of the conditions were generated using GCR F_{crit} indices (Harris, 1975, page 120) for significance testing. The mean scores of the conditions were different for all six dependent variables (fact recall, $F = 81.07$; concept comprehension, $F = 111.94$; general synthesis skill, $F = 28.06$; social problem solving, $F = 24.58$; interest, $F = 24.23$; involvement, $F = 27.51$; in all cases $df = 3/149$ and $p < .01$ for $F_{crit} = 9.67$). In addition to this univariate analysis, pairwise comparisons of the mean scores of the conditions were conducted using Roy and Bose (1953) contrasts. The results of the pairwise comparisons are displayed in Table 2.

Retention test analyses for above average readers

The data were submitted to a simple MANOVA. The mean composite scores of the conditions were different (GCR $\theta = .687$, $s = 3$, $m = 1$, $n = 71$, $p < .0001$; Wilks' $\Lambda = .151$, $F = 21.52$, $df = 18/407$, $p < .0001$). The composite associated with the condition's effect was (fact recall + concept comprehension - general synthesis skill). Again, a tentative interpretation of the composite as a discrepancy between lower and higher taxonomic level skills could be posed.

Univariate analyses using GCR F_{crit} indices (Harris, 1975) indicated the conditions were different for all dependent variables (fact recall, $F = 49.69$; concept comprehension, $F = 42.94$; general synthesis skill, $F = 16.63$; social problem solving, $F = 21.33$; interest, $F = 20.67$; involvement, $F = 24.66$; in all cases $df = 3/149$ and $p < .01$ for $F_{crit} = 9.67$). Pairwise comparisons of the conditions' means were conducted using Roy and Bose (1953) contrasts. The results of the pairwise comparisons are displayed in Table 3.

Analyses for Below Average Readers

The condition X test occasion data of the below average readers were submitted to a 4 x 2 repeated measures MANOVA. A condition X test occasion interactive effect was identified (GCR θ = .599, s = 3, m = 1, n = 70.5, $p < .0001$; Wilk's Λ = .362, F = 9.74, df = 18/405, $p < .0001$). As with the prior interactions, the condition X test interaction was viewed as empirical justification for separate analyses of the post and retention test data.

Post-test analyses for the below average readers

The data were submitted to a simple MANOVA. The mean composite scores of the subjects in the four conditions were different (GCR θ = .613, s = 3, m = 1, n = 70.5, $p < .0001$; Wilks' Λ = .266, F = 13.42, df = 18/405, $p < .0001$). The composite associated with the effect of condition was (2* fact recall - general synthesis skill). Again, the composite appeared to be focused upon a discrepancy between lower and higher taxonomic level skills.

Univariate analyses using GCR F_{crit} indices (Harris, 1975) indicated the conditions differed on fact recall (F = 60.63, df = 3/148 and $p < .01$ for F_{crit} = 9.68), concept comprehension (F = 28.63, df = 3/148 and $p < .01$ for F_{crit} = 9.68), and general synthesis skill (F = 8.67, df = 3/148 and $p < .05$ for F_{crit} = 7.70). For the other dependent variables, the differences in the conditions' mean scores failed to exceed chance expectation. The results of pairwise comparisons of the conditions' mean fact recall, concept comprehension, and general synthesis skill scores are displayed in Table 4.

Retention test analyses for the below average readers

The data were submitted to a simple MANOVA. The mean composite scores of the subjects in the four conditions were different (GCR θ = .447, s = 3, m = 1, n = 70.5, $p < .0001$; Wilks' Λ = .400, F = 8.61, df = 18/405, $p < .0001$). The composite associated with the effect of condition was (fact recall + concept comprehension - general synthesis skill). Again, as with all other cases, the composite appeared to be focused upon a discrepancy between lower and higher taxonomic level skills.

Univariate analyses indicated the conditions differed on fact recall and concept comprehension (F = 21.58, F = 20.19, respectively, with df = 3/148, $p < .01$ for F_{crit} = 9.68). Roy and Bose (1953) comparisons indicated the control and expository groups' mean fact recall scores were different (GCR θ = .200, s = 3, m = 1, n = 70.5, $p < .0005$). The control and expository groups' mean concept comprehension scores were different (GCR θ = .244, s = 3, m = 1, n = 70.5, $p < .0001$). Finally, comparison of the control and guided inquiry groups' mean fact recall and concept comprehension scores resulted in differences (GCR θ = .286 and GCR θ = .216, respectively, s = 3, m = 1, n = 70.5, $p < .0001$).

Conclusions

The results indicated the impact of the conditions was dependent upon the reading ability of the students. However, the major differences attributable to conditions appeared to be a discrepancy in lower and higher taxonomic level skills. For the above average readers, the conditions displayed relatively long term effects, with the most stable differences confined to the lower level cognitive skills. For the below average readers, the conditions displayed relatively short term effects. In fact the conditions failed to display any effects on the affective or higher level cognitive structures of the below average readers.

Summary

The repeated measures MANOVA technique includes a comprehensive, structured hypothesis set. The empirical justification for separate analyses, as found in the example, is a major benefit not present with most other techniques. In addition, the composite structures frequently provide the analyst with serendipitous "constructs" revealing informative sources of the impact of independent variables. Clearly, this technique is a viable analytic tool that may be applied to data that has previously resisted pristine analysis.

The sophistication required of the analyst using repeated measures MANOVA is not very much greater than for repeated measures ANOVA or higher order MANOVA. Indeed, repeated measures MANOVA should be useful and usable for the majority of educational researchers.

References

Carroll, C. D. and Klimowicz, E. L. RMV: A FORTRAN program for repeated measures multivariate analysis of variance. In Carroll, C. D. Education Statistics Package, User's Manual. School of Education, University of California, Riverside, 1975.

Elashoff, J. Analysis of covariance, a delicate instrument. AERJ, 1969, 6, 383-401.

Harris, R. J. A Primer of Multivariate Statistics. New York: Academic Press, 1975.

Michaelis, J. U. Preparation of teaching guides and materials on Asian countries for use in grades 1-12. University of California, Berkeley, 1968 (OE-5-10-040).

Morrison, D. F. Multivariate Statistical Methods. New York; McGraw-Hill, 1967.

Roy, S. N. and Bose, R. C. Simultaneous confidence interval estimation. Annals of Mathematical Statistics, 1953, 24, 513-516.

Table 1
Means ($\bar{X}$), Standard Deviations (SD), and Sample Sizes (N) for the Condition x Reading Level x Test Occasion Arrangement of Subjects' Scores on the Six Dependent Variables

	Control		Expository		Guided Inquiry		Open Inquiry	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Above Average Readers:								
N	38		37		40		38	
Fact Recall:								
Post	9.18	3.53	26.24	3.72	23.85	6.24	19.05	6.36
Retention	8.05	2.97	21.84	4.36	19.87	6.35	15.45	6.73
Concept Comprehension:								
Post	7.47	3.01	26.65	2.64	22.57	6.69	18.97	5.43
Retention	8.63	3.62	21.37	5.68	18.92	6.12	16.66	4.93
Social Problem Solving:								
Post	11.45	4.15	13.43	5.54	20.50	3.88	15.42	5.83
Retention	10.79	3.95	12.70	5.60	18.72	3.72	14.92	4.99
General Synthesis Skill:								
Post	12.47	4.23	16.38	5.49	20.00	4.59	22.18	5.46
Retention	13.59	4.29	12.73	4.57	17.55	6.76	20.34	4.80
Interest:								
Post	25.89	12.23	36.51	10.22	45.00	6.59	34.66	10.03
Retention	24.84	12.58	34.24	11.00	43.45	5.55	33.08	11.52
Involvement:								
Post	25.97	12.00	30.38	8.69	43.97	6.38	34.97	8.86
Retention	27.50	10.47	26.59	9.56	42.85	6.04	33.32	10.91
Below Average Readers:								
N	33		37		46		36	
Fact Recall:								
Post	2.84	2.11	18.68	6.40	15.00	5.59	11.89	4.98
Retention	3.06	2.37	10.03	5.47	11.89	5.55	8.19	5.33
Concept Comprehension:								
Post	3.36	2.60	15.51	7.21	12.85	6.66	7.86	6.01
Retention	2.39	3.08	10.78	6.10	10.15	5.44	5.91	5.62
Social Problem Solving:								
Post	4.39	4.64	6.14	5.50	7.71	6.44	5.39	4.80
Retention	4.70	5.29	6.03	4.73	6.93	6.72	5.11	5.01
General Synthesis Skill:								
Post	6.33	5.51	7.03	4.94	11.89	5.87	7.83	5.44
Retention	6.42	5.97	5.27	4.55	10.07	5.42	6.78	6.16
Interest:								
Post	14.82	11.31	25.92	12.08	22.85	12.13	24.53	11.82
Retention	18.12	10.95	25.10	13.87	21.86	10.75	23.36	11.04
Involvement:								
Post	16.91	9.90	23.05	10.51	25.52	12.47	20.08	9.47
Retention	14.82	11.27	18.57	11.75	24.33	13.41	21.14	12.08

Table 2

Results of Roy and Bose (1953) Comparisons 1/ of Conditions for Post-Test Scores of the Above Average Readers

Condition 2/ Pair	Fact Recall	Concept Comprehension	General Synthesis Skill	Social Problem Solving	Interest	Involvement
C, E	$\theta = .583$ $p < .0001$	$\theta = .672$ $p < .0001$	n.s.	n.s.	$\theta = .128$ $p < .05$	n.s.
C, GI	$\theta = .508$ $p < .0001$	$\theta = .560$ $p < .0001$	$\theta = .227$ $p < .0001$	$\theta = .304$ $p < .0001$	$\theta = .322$ $p < .0001$	$\theta = .330$ $p < .0001$
C, OI	$\theta = .319$ $p < .0001$	$\theta = .424$ $p < .0001$	$\theta = .329$ $p < .0001$	n.s.	n.s.	n.s.
E, GI	n.s.	n.s.	n.s.	$\theta = .210$ $p < .0001$	n.s.	$\theta = .220$ $p < .0001$
E, OI	$\theta = .199$ $p < .001$	$\theta = .247$ $p < .0001$	$\theta = .149$ $p < .01$	n.s.	n.s.	n.s.
GI, OI	n.s.	n.s.	n.s.	n.s.	$\theta = .122$ $p < .05$	n.s.

1/ s = 3, m = 1, n = 71 for all comparisons using GCR (Harris, 1975).
2/ C = Control, E = Expository, GI = Guided Inquiry, OI = Open Inquiry.

Table 3

Results of Roy and Bose (1953) Comparisons 1/ of Conditions for Retention Scores of Above Average Readers

Condition 2/ Pair	Fact Recall	Concept Comprehension	General Synthesis Skill	Social Problem Solving	Interest	Involvement
C, E	$\theta = .460$ $p < .0001$	$\theta = .436$ $p < .0001$	n.s.	n.s.	n.s.	n.s.
C, GI	$\theta = .386$ $p < .0001$	$\theta = .335$ $p < .0001$	n.s.	$\theta = .275$ $p < .0001$	$\theta = .288$ $p < .0001$	$\theta = .225$ $p < .0001$
C, OI	$\theta = .197$ $p < .0005$	$\theta = .235$ $p < .0001$	$\theta = .161$ $p < .005$	n.s.	n.s.	n.s.
E, GI	n.s.	n.s.	n.s.	$\theta = .180$ $p < .005$	n.s.	$\theta = .277$ $p < .0001$
E, OI	$\theta = .155$ $p < .01$	n.s.	$\theta = .214$ $p < .0001$	n.s.	n.s.	n.s.
GI, OI	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

1/ s = 3, m = 1, n = 71 for all comparisons using GCR (Harris, 1975).
2/ C = Control, E = Expository, GI = Guided Inquiry, OI = Open Inquiry.

Table 4

Results of Roy and Bose (1953) Comparisons 1/ of Conditions for Post-Test Scores of Below Average Readers

Condition 2/ Pair	Fact Recall	Concept Comprehension	General Synthesis Skill
C, E	$\theta = .546$ $p < .0001$	$\theta = .341$ $p < .0001$	n.s.
C, GI	$\theta = .415$ $p < .0001$	$\theta = .239$ $p < .0001$	n.s.
C, OI	$\theta = .282$ $p < .0001$	n.s.	n.s.
E, GI	n.s.	n.s.	n.s.
E, OI	$\theta = .181$ $p < .005$	$\theta = .170$ $p < .005$	n.s.
GI, OI	n.s.	n.s.	n.s.

1/ s = 3, m = 1, n = 70.5 for all comparisons using GCR (Harris, 1975).
2/ C = Control, E = Expository, GI = Guided Inquiry, OI = Open Inquiry.